

# Indirect and mixed treatment comparisons in arthritis research

Anthony E. Ades<sup>1</sup>, Jason Madan<sup>1</sup> and Nicky J. Welton<sup>1</sup>

## Abstract

Evidence for the efficacy of biologic therapies in inflammatory arthritis comes overwhelmingly from placebo-controlled trials. Increasingly, however, authorities responsible for purchasing and re-imbursement have tried to determine whether there are differences between these powerful new therapies, which would lead them to recommend some in preference to others, either on grounds of efficacy or cost-effectiveness. In the absence of head-to-head trial comparisons, indirect comparisons may be used. Furthermore, network meta-analysis, also known as mixed treatment comparisons can combine information from trials in a connected network. These methods allow inferences about head-to-head comparisons even when there is little or no head-to-head evidence, which has caused some concern. In this article we briefly review these methodologies and describe results from recent applications to inflammatory arthritis in the clinical literature. We then focus on how the methodologies are used in decision making, taking as an illustration some recent technology appraisals conducted by the National Institute for Health and Clinical Excellence in the UK. We conclude that, in practice, the key decisions have been based on results from placebo-controlled trials.

**Key words:** Indirect comparisons, Mixed treatment comparison, Network meta-analysis, Biologic therapies.

## Introduction

It is widely accepted that randomized trials provide the best evidence to inform treatment decisions, whether these decisions are to be based on efficacy alone or on efficacy and cost-effectiveness. In the case of treatments for RA and PsA, new and powerful agents, the anti-tumour necrosis factor (TNF) inhibitors, have come into increasing use in recent years. Their efficacy against placebo is well established at several points in the treatment pathway. Increasingly, however, decision makers have wanted to discriminate between the alternative products and select the best.

Indirect comparisons (ICs) [1] allow one to draw inferences about the relative effect of treatments B and C from data in A vs B and A vs C trials. More generally, we may have data from AB, AC and BC trials, and indeed, AD, BD and so on. Statistical methods have been available since 1994 for creating a basis for multiple

treatment comparison from an evidence base that consists of a series of pair-wise comparisons [2–4]. This is usually referred to as network meta-analysis (NWMA) [4] or mixed treatment comparisons (MTCs). Both ICs and NWMA are, in effect, extensions of standard pair-wise meta-analysis [5, 6]. The growing importance of health technology assessment, based on a cost-effectiveness analysis (CEA), has been accompanied by a rapid adoption of these methods, and recently a series of papers have produced a formulation that can apply to networks consisting of any number of trials on any number of treatments [7–10].

The popularity of these methods in a decision-making context is not a coincidence. Decision making in the presence of multiple treatment options requires an internally consistent and coherent model of relative treatment effects. This can be conceptualized as follows. If we consider three treatments X, Y and Z, then coherent means that the true effect of treatment Z relative to treatment X,  $d_{XZ}$ , must be equal to the sum of the effects of Y relative to X and Z relative to Y:  $d_{XZ} = d_{XY} + d_{YZ}$ . This is shown in Fig. 1.

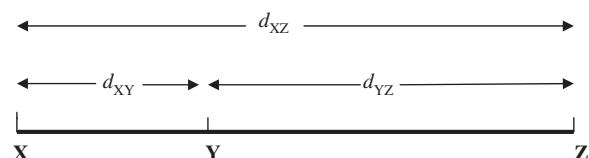
Another way to understand this, perhaps, is that if Peter is 3 inches taller than Paul, and Paul is 2 inches taller than Mary, then Peter must be 5 inches taller than Mary. Clearly, there is nothing controversial about this simple relationship. The point is, however, that if we were to

<sup>1</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK.

Submitted 2 February 2011; revised version accepted 7 June 2011.

Correspondence to: Anthony E. Ades, School of Social and Community Medicine, University of Bristol, Bristol, UK.  
E-mail: t.ades@bristol.ac.uk

**Fig. 1** Schematic of coherent treatment effects, in which the relative effects of treatments X, Y and Z must have the property  $d_{XY} + d_{YZ} = d_{XZ}$ .



abandon the concept of coherence, which allows us to deduce the three height differences from any two, then we would not have a rational way of deciding who is the tallest of these three people. In the same way, to decide which of the three treatments is best we need a form of evidence synthesis that provides us with a coherent model of treatment differences. We also want to be sure, of course, that the data fit the model.

ICs and NWMA are generalizations of pair-wise meta-analysis in which, rather than conducting separate meta-analyses of the XY, XZ and YZ trials, a single synthesis is carried out in which all three relative treatment effects are estimated, subject to the constraint that the estimates must be coherent in the sense above. This is the reason why they feature inevitably in decision making when there are more than two options.

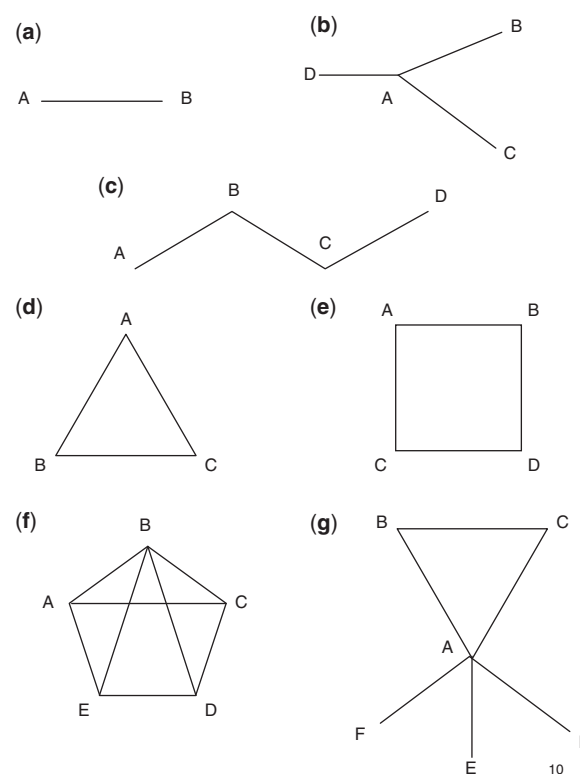
In this article we briefly review the properties of these statistical methods and then examine applications to RA and PsA in the clinical literature. Readers may consult general texts on meta-analysis [11] and previous methodology work on ICs, MTCs [8–10, 12] as well as the rheumatology applications cited below. The main focus of this article is on the way that ICs and NWMA have influenced decision making. To illustrate this we look specifically at submissions to the National Institute for Health and Clinical Excellence (NICE) in the UK.

## Networks of randomized trials

Figure 2 shows a range of connected networks. Each edge or line connects two treatments, and the existence of a line means that these two treatments have been compared in a randomized trial. The simplest network is, of course, the pair-wise meta-analysis [Fig. 2(a)]. The star network [Fig. 2(b)] and the chain [Fig. 2(c)] are both structures that allow ICs to be made. In the star structure, all trials share a common comparator, A (e.g. MTX plus placebo), to which B, C, D have been compared, and where these might represent MTX + etanercept, MTX + infliximab and MTX + adalimumab. The network, and the statistical model, allows us to use the AB, AC and AD evidence to form conclusions about BC, BD and CD relative effects. Similarly, the chain structure allows inference on AD effects, based on AB, BC and CD trials—though this structure is not commonly seen in arthritis research.

At the next level of complexity are the triangle or square structures [Fig. 2(d) and 2(e)]. These have the additional feature that the direct evidence in any one edge, say AB, can be compared with the indirect evidence formed from

**Fig. 2** An illustration of possible structures for networks of evidence.



the AC and BC edges, or the AC, CD and BD edges. Finally, there are networks of general complexity [Fig. 2(f) and 2(g)], which require specially tailored regression models. At the time of writing, a typical network structure in the inflammatory arthritis field would consist mainly of a star structure, with placebo, or placebo + MTX as the common comparator, and with a limited number of head-to-head trials providing direct comparisons of different biologics. Figure 2(g) depicts a network of this sort, with a single head-to-head comparison between the active biologic treatments.

An important point is that the uncertainty in an IC is always greater than the uncertainty in the direct comparisons from which it is composed. The variances add along the chain. Inference concerning the AD comparison in network (c) in Fig. 2 is based on three links, and has the summed variance of each.

Although these are very powerful synthesis methods, which automatically generate coherent comparisons in the sense discussed earlier, the key assumptions behind them are only subtly different from those in pair-wise meta-analysis. Pair-wise meta-analysis assumes that each AB trial estimates the same (fixed) or similar (random) AB effect, and that each AC trial estimates the same or similar AC effect. To obtain a valid network analysis with the all-important coherence property, one only needs to assume that the scope of these assumptions extends beyond the set of trials in which the specific

contrasts are made to include all the trials in the network. Thus the true AB effect would be the same in the AC, AD and even in the CD trials—if only treatments A and B had been included in them.

This is an assumption, however, that deserves careful scrutiny. It is known, for example, that response to RA treatment depends on time since diagnosis (see below). If the relative treatment effects also depend on time since diagnosis, then combining trials with different average times since diagnosis will produce uninterpretable results, unless time since diagnosis is allowed for in the analysis using subgroup analyses or meta-regression [13]. Although this is true of both pair-wise and network synthesis, comparisons between similar products based on small amounts of indirect data are clearly especially vulnerable. For example, in an IC between two biologics with placebo as a common comparator, apparent differences could be entirely due to these confounders. Head-to-head comparisons between biologics will, in contrast, be perfectly sound, as the effect of variables such as disease severity or duration of illness can be expected to be the same on both active treatments, leaving their relative effects unaltered.

## Results from published network meta-analyses in inflammatory arthritis

NWMA is being applied with increasing frequency to RA and PsA in a decision-making context, but results have also been appearing in the clinical literature. A study by Nixon *et al.* [14] compares biologics against placebo, and biologics + MTX against placebo + MTX, with 6-month ACR-50 status as the outcome. Thirteen trials were identified, involving various doses of the TNF- $\alpha$  inhibitors etanercept, infliximab and adalimumab, and the IL-1 inhibitor anakinra. This body of evidence not only shows the dramatic effects of biologics, but uses meta-regression to illustrate convincingly that their relative effects increase with time since diagnosis. When this is not taken into account, there appear to be substantial differences between the biologics, but when it is allowed for the anti-TNF- $\alpha$  drugs emerge as very similar but superior to anakinra. The effect of disease duration is to increase the odds ratio advantage of biologics by a factor of 1.13 per year.

A second study by Singh *et al.* [15], based on an overview of Cochrane reviews, takes an even wider perspective. It includes a further eight studies involving etanercept, infliximab, adalimumab or anakinra, and also seven studies involving the anti-CD-28 therapy abatecept, and three on the anti-B-cell drug rituximab. Like the Nixon *et al.* article [14], the structure is a star network. (As an aside, this does not necessarily mean that there are no head-to-head comparisons between biologics, only that there are no Cochrane systematic reviews of head-to-head comparisons.) But unlike the Nixon *et al.* study [14], the synthesis pools information over a very broad range of patient groups, varying not only in disease severity and duration, but also in whether patients had failed

on a previous biologic therapy, and trials involving dual-biologic therapy. In the systematic review, a series of subgroup analyses are presented, for example: concomitant MTX or not; early, established, late RA; anti-TNF inhibitor or not; previous failure on biologics, non-biologic DMARDs or neither; and so on, with each analysis pooling information over all the other factors, including the treatments. These analyses are very difficult to interpret. The authors again report that the efficacy of biologics increases with disease duration, and also that they are similarly effective in patients who have failed on biologics, or failed on previous non-biologic DMARDs, or neither; and that all biologics were similarly effective except for anakinra. However, each of these subgroup analyses is confounded by the other variables. It is not possible to draw the conclusion that the relative effects of biologics are the same in each subgroup, as no tests of interaction are presented and there are insufficient data to detect such interactions. In addition, the different classes of biologics have somewhat different licensed indications, making it inevitable that treatment comparisons are confounded by key patient characteristics.

## NWMA of RA and PsA therapies in decision making at NICE

Experts have long recognized that, unlike reviews aimed at literature summary, in a decision-making context, systematic review and evidence synthesis has to be more focused [16]. A decision to recommend a treatment, whether based on efficacy or cost-effectiveness, is a decision that relates to specific treatments, at specific doses with specified concomitant therapy, in specific classes of patients. The analyst cannot, therefore, present a series of analyses and leave the reader to decide whether they are credible, but must adopt a series of positions that can be shown are supported by evidence.

The series of recent decisions at NICE in the UK represent a valuable case study not only because network meta-analyses are frequently presented in submissions to NICE, but because the transparency of procedure and accessibility to the key documents and decisions gives us a unique insight into the role of indirect data in decision making.

Health technology evaluation at NICE is undertaken on the basis of both comparative efficacy and cost-effectiveness. A methods guide, revised in 2008 [17], sets out the preferred methodology, and this is incremental probabilistic CEA. Decisions are based on expected net benefit [18], which is monetarized (lifetime) health gain where quality-adjusted life years are valued at £20 000–£30 000/year minus lifetime costs to the National Health Service (NHS). NICE's position on ICs is that direct head-to-head evidence is preferred, but ICs are accepted when direct evidence is lacking. Analyses pooling direct and indirect data, i.e. MTCs, can be presented as subsidiary analyses [17]. However, where a choice must be made between three or more interventions, the logic of

incremental CEA requires a coherent model in the sense shown in Fig. 1, and it is recognized that network synthesis is the only way this can be delivered. NICE appraisal committees, however, have considerable discretion to determine which data and which form of analysis are most appropriate in each case, although they are careful to provide a reasoned argument, particularly as they may eventually be asked to explain the basis for their decisions in the courts [19].

We begin with PsA, where three treatments—etanercept, infliximab and adalimumab—had already been approved for use in patients who had failed on two conventional DMARDs, following separate submissions by their manufacturers to three separate single technology appraisals. In 2010 NICE undertook a review of these decisions, later published as TA199 [20]. In this multiple technology assessment, one of the university groups contracted for this work produced an assessment document that formed the main basis for NICE's decision. The key evidence was from six randomized controlled trials, two on each drug, comparing the biologic therapy with placebo, with Psoriatic Arthritis Response Criteria, ACR and Psoriasis Area Severity Index outcomes. This represents a three-prong star network for an IC analysis, exactly as shown in Fig. 2(b).

In general, all three products were highly effective and highly cost effective compared with placebo. Infliximab was the most effective, but differences were not statistically significant. Infliximab treatment was also the most costly. The formal incremental CEA showed that etanercept was the cost-effective option: a strict application of the rules of CEA would therefore have led the committee to recommend etanercept, but not the other two TNF- $\alpha$  inhibitors. However, the committee followed a different line of reasoning. First, they acknowledged that the IC did not suggest that there was any material difference in efficacy between the therapies, and that this accorded with what one might assume *a priori*. Second, each was clearly cost effective compared with no treatment. However, there were cost differences, although their precise extent depended on the prices that individual purchasing authorities were able to negotiate. NICE therefore recommended all three products for use in the NHS, in effect confirming the three original decisions, but with the key proviso that authorities should use the least costly. The important point to note here is that in the end the evidence driving the decision was in fact the original trial comparisons with placebo. The ICs, because of their higher variance, and probably the lack of true differences, fail to provide evidence that convincingly rules any of the products out. But there may also be a political aspect to this. A strict application of CEA, which would have ruled out two of the three products, might have sparked appeals from the manufacturers. NICE would then have to justify ruling against two products in the absence of convincing evidence that they were less effective. The way the decision was formulated appears to allow NICE to avoid being put in this position, while still securing an economically reasonable outcome for the NHS.

A somewhat different example is furnished by the assessment of anakinra in RA [21]. The critical trials were those of anakinra + MTX vs placebo + MTX, and anakinra vs placebo. In this case the ICs carried out by the assessment group, in common with those reported in the clinical literature [15, 22], suggested that anakinra was inferior to other biologics. However, the decision that was made to not recommend anakinra was not related to the IC, but was simply made on the basis that, at the manufacturer's price and based on the assessment group's model, anakinra was not cost effective compared with no anakinra with or without MTX as concomitant treatment.

In 2007 NICE recommended etanercept, infliximab and adalimumab as dual therapy with MTX, or alone in patients unable to take MTX [23], for RA patients who had failed on conventional DMARDs. Three years later the manufacturers of certolizumab, another anti-TNF- $\alpha$  treatment, submitted their product for consideration for these indications. They assembled an evidence base consisting of the certolizumab trials and earlier comparisons of the anti-TNF- $\alpha$  drugs and placebo [24]. The incremental CEA model presented by the manufacturers showed that certolizumab was the cost-effective option in both cases. The appraisal committee at NICE did not agree that certolizumab was the cost-effective option. However, arguing that there was no basis for believing that it was any less or any more effective than the other anti-TNF- $\alpha$  products, they recommended certolizumab for use in RA alongside the competitor drugs, on the basis that it was, like its competitors, highly cost effective against no treatment, based on the comparisons with placebo. Once again, the network synthesis, which was finally based on a network with a structure like that in Fig. 2(g), allowed NICE to consider the efficacy and cost-effectiveness of certolizumab in the context of all the competitor therapies, but the decisive information supporting the decision was once again the direct evidence from placebo-controlled trials.

## Conclusion

Although fears are often expressed about the reliability of ICs, their influence in formal decision making is automatically limited by their relative uncertainty. It would obviously be of interest to conduct a more comprehensive survey of the impact of indirect evidence on decision making both at NICE and in other re-imbursement authorities, and for a wider range of conditions. However, the evidence presented here on PsA and RA suggests that biologic therapies have been approved for use on the basis of comparison with conventional therapy, and products have not been ruled out by comparison with competitor products—except on the basis of price.

It is interesting to note that although calls are often made for more head-to-head comparisons between biologics, given that these products are generally approved for use on the basis of placebo-controlled trials, the effect of additional head-to-head evidence would only be to

rule out one or more biologics that would otherwise be in use.

#### Rheumatology key messages

- Inferences can be made about the comparative efficacy of treatments even when head-to-head evidence is sparse.
- ICs have not played a significant role in NICE decision-making on biologics.

**Supplement:** This paper forms part of the supplement 'Biologic therapies in inflammatory joint diseases: models, evidence and decision making.' This supplement was supported by unrestricted funding from Arthritis Research UK.

**Disclosure statement:** The authors have declared no conflicts of interest.

## References

- 1 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.
- 2 Gleser L, Olkin I. Meta-analysis for 2X2 tables with multiple treatment groups. In: Stangl DK, Berry DA, eds. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker, 2000.
- 3 Higgins JPT, Whitehead J. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15: 2733-49.
- 4 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313-24.
- 5 Higgins JPT and Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell, 2008.
- 6 Sutton AJ, Ades AE, Cooper N, Abrams KR. Use of mixed and indirect comparisons in technology assessment. *Pharmacoeconomics* 2008;26:753-67.
- 7 Lu G, Ades A. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101: 447-59.
- 8 Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials 2011, Available from <http://www.nicedsu.org.uk>.
- 9 Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br Med J* 2005;331:897-900.
- 10 Ades AE, Sculpher M, Sutton A *et al*. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006;24:1-19.
- 11 Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. London: Wiley, 2000.
- 12 Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23: 3105-24.
- 13 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21:1559-74.
- 14 Nixon R, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* 2007;26:1237-54.
- 15 Singh JA, Christensen R, Wells GA *et al*. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *Can Med Assoc J* 2009;181:787-96.
- 16 Eccles M, Freemantle N, Mason J. Using systematic reviews in clinical guideline development. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group, 2008.
- 17 National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal*. London: NICE, 2003.
- 18 Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analyses. *Med Decis Making* 1998;18:S68-80.
- 19 Syrett K. NICE and judicial review: enforcing 'accountability for reasonableness' through the courts? *Med Law Rev* 2008;16:127-40.
- 20 National Institute for Health and Clinical Excellence. *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis - Technology Appraisal TA199*. London: NICE, 2010.
- 21 National Institute for Health and Clinical Excellence. *The clinical effectiveness and cost-effectiveness of anakinra for rheumatoid arthritis - Technology Appraisal TA72*. London: NICE, 2003.
- 22 Nixon R, Bansback N, Brennan A. The efficacy of inhibiting tumour necrosis factor alpha and interleukin 1 in patients with rheumatoid arthritis: a meta-analysis and adjusted indirect comparisons. *Rheumatology* 2007;46:1140-7.
- 23 National Institute for Health and Clinical Excellence. *Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis - Technology Appraisal TA130*. London: NICE, 2007.
- 24 National Institute for Health and Clinical Excellence. *Certolizumab pegol for the treatment of rheumatoid arthritis - Technology Appraisal TA186*. London: NICE, 2010.